How Pressing ‘ENTER’ Gets You to The Hague

Lisa M. Cohen, Lorenz Rubner 2022-03-02T08:00:03

The offensive use of cyber capabilities is well documented. Most recently, the Russian invasion into Ukraine has been preceded and accompanied by multiple cyber operations both from State actors and civilian collectives. The employment of offensive cyber capabilities as part of international armed conflicts has now become an evident reality – and it raises the question: Can the execution of cyber operations trigger international individual criminal responsibility under Art. 8(2) of the Rome Statute (‘RS’) of the International Criminal Court (‘ICC’)? The short answer is yes. While the inclusion of cyber-specific war crimes within the RS has received little to no traction (see here), Rule 84 of the Tallinn Manual 2.0 plainly states that “cyber operations may amount to war crimes”. In November 2021, the Council of Advisers on the Application of the Rome Statute to Cyberwarfare (‘Council’), a group of international legal and technical experts, examined the details of *how* the RS’s core crimes can be applied to cyberwarfare but only briefly linked cyber operations to *specific* crimes under Art. 8(2) RS.

This post aims to provide an overview of how different categories of cyber operations may fall within the RS’s framework for war crimes. To this end, it will distinguish between cyber operations with direct or indirect injurious or damaging effects (also dubbed kinetic cyber or cyber-physical operations), and those operations which do not entail any such physical effect. For the purpose of this analysis, we assume that an international armed conflict exists, attribution is possible, and the cyber operations are not targeted at military objectives.

Pulling the Virtual Trigger: Injury or Damage as Primary Effects

According to the Tallinn Manual 2.0, direct effects are the “immediate, first order consequences […], unaltered by intervening events or mechanisms” (p. 472). Imagine, for example, a malware causing a pacemaker to halt its functioning, foreseeably leading to the person’s immediate death; or consider State A attacking a uranium enrichment facility in the enemy’s territory with malware “forc[ing] a change in the centrifuge’s rotor speed [and] inducing excessive vibrations or distortions that would destroy the centrifuge” (cf. Stuxnet) or causing a severe explosion.

It is important to highlight that the physical effects a cyber activity can produce will never be as *immediate* as with conventional weapons. Cyber operations’ primary effects will first and foremost be a change in or overwhelming of the target system (software). This, however, must not obscure the fact that – much like the pulling of a trigger initially merely releases the firing pin, which in turn initiates the next chain of physical (injurious or damaging) reactions – a cyber weapon’s immediate effects are not the end of a causal reaction that is as foreseeable and intended as the entry of a

bullet into the envisaged target. Consequently, those physical injurious or damaging effects should be considered the cyber operation's *direct* effect.

Due to the apparent similarities, such cyber operations should be treated no differently than traditional kinetic attacks when considering the application of the RS. This applies in particular to the term "attack" which is frequently used in Art. 8(2) RS and serves as the entry point for the application of many core international humanitarian law (IHL) principles protected by the RS (p. 37 ff. of the Report). Art. 49(1) Additional Protocol I to the Geneva Conventions (API) defines an attack as "acts of violence against the adversary, whether in offense or defense". It is well settled in IHL that attacks are not "limited to activities that release kinetic force" (p. 415) but rather also include non-violent acts (e.g., pressing "ENTER") with violent effects. The described cyber operations would therefore be war crimes of willful killing (Art. 8(2)(a)(i) RS) or attacks against civilian objects (Art. 8(2)(b)(ii) RS). As isolated incidents, however, both operations will likely not cross the gravity threshold under Art. 17(d) RS. ICC investigations would only be admissible with particularly severe consequences or combined with other cyber or traditional military operations.

Further Down the Chain of Causality: Injury or Damage as Secondary Effects

The Tallinn Manual holds that indirect effects of a cyberattack include "the delayed and/or displaced […] higher-order consequences […] created through intermediate events or mechanism" (p. 472). Consider the following cases: State A employs a malware targeted at a water treatment facility in the enemy's territory, changing the chemical composition in such a way that the water becomes toxic (cf. the incident in Florida), or one which halts the functioning of an enemy cities' only natural gas pipeline with the intent of depriving the population of fuel. The civilian population loses its only source of power, thereby foreseeably resulting in death due to harsh winter conditions (cf. the Colonial Pipeline hack).

The primary physical effects – the poisoning of the water and blocking of gas – can equally be achieved by non-cyber means. As such, those cyber operations should be treated identically to their kinetic counterparts. Even if the possible secondary physical effects – death of civilians – were not to materialize, such cyber operations can constitute war crimes by virtue of their primary effects alone, cf. "employing poison" (Art. 8(2)(b)(xvii) RS) and "depriving civilians of objects indispensable to their survival" (Art. 8(2)(b)(xxv) RS). Rendering a gas pipeline useless may also in itself be an act of destruction or extensive damage of property (e.g., Art. 8(2)(a)(iv), (b)(xiii) RS). While the drafters of the Geneva Conventions and the RS clearly had more permanent and physical destruction in mind (para. 3120), the ordinary meaning may be construed to include the non-permanent loss of functionality (cf. pp. 417 f.).

Focusing on the secondary physical effects in the abovementioned examples, death of civilians, these too may clearly constitute war crimes as long as the cyber operation is a "substantial cause of the death of the victim" (para. 296). The ICC has not elaborated much further on the element of causation but generally seems to apply a simple but/for test (i.e., *conditio sine qua non*), which allows for significant freedom to consider a range of actions as causation, including cyberattacks.

Intangible but Not Inconsequential: Non-Physical Effects

Cyber operations causing (only) non-physical effects will likely be most problematic when it comes to the question of whether they can amount to war crimes. Consider the following example: State A targets the enemy state's largest transportation company with ransomware, encrypting most of its operating data and resulting in prolonged business interruption and immense financial losses (cf. the Maersk incident caused by NotPetya). Imagine, as a second example, State A gaining access to the network of a large corporation based within the enemy's territory, copying business secrets and confidential employee data and publicizing parts thereof and using other parts for coercion (cf. Sony Pictures). Both these examples bear consequences that are by no means confined to the cyber world. However, as opposed to the examples with (in)direct physical effects pictured above, cyber operations categorized as *non-physical* do not carry injury or destruction as primary or secondary effects.

According to Germany's cyber position paper, "[t]he occurrence of physical damage, injury or death to persons or damage or destruction to objects comparable to effects of conventional weapons is *not* required for an attack in the sense of art. 49 para. 1 [API]", a view shared by the Council. Their Report, however, highlights that Art. 49(1) API establishes "violence" as the essential criteria to differentiate attacks from other military operations. "Cyber operations that are by their nature non-violent, such as espionage or psychological operations, cannot be considered attacks" (p. 38). While the Report regards "disrupting or halting the functions of a State's critical infrastructure or jamming military capabilities" (p. 38) to be a violent attack, even if there is no physical destruction, it is much more questionable whether the disruption of a private company or the publication of data can be considered "violent". The "mere intrusion into foreign networks and the copying of data" is specifically excluded from attacks by the German analysis (p. 8).

While the definition of "attack" is highly relevant to many war crimes (see above), the most disputed issue of non-physical cyber operations lies outside of this element: Can data be construed as objects? While the Tallinn Manual 2.0 rejected this view (Rule 100), the Report now explicitly diverged from this analysis (p. 39), further adding to a contested debate (see e.g., here, here, and here). This dispute extends to the definition of property under Art. 8(2)(a)(iv), (b)(xiii), and (e)(xii) RS.

Returning to the examples on cyber operations with non-physical effects, the encryption of data may constitute a war crime if it also disables critical infrastructure. Under an extensive interpretation of the RS, the encryption of a private company's data leading (only) to financial losses may also be a war crime if data is considered an object. In this case, the targeting of civilian data must either be considered an attack, i.e., an act of violence or be handled in such a way as to be considered destruction, appropriation, or seizure of property. This aligns with the view of the Council, which advocated for the alteration or deletion of civilian medical data to be considered a war crime (p. 47). The copying of data and the mere (threat of) publication for coercion, however, will likely not fall under the RS unless it is considered a "violent" attack.

Cyber War Crimes: New Norms Needed?

But should such cyber operations be considered war crimes? The public debate is often quick to claim *accountability gaps*. But a cautious approach to the demand for an amendment to the RS is warranted. Clearly, unauthorized access to personal data violates the right to privacy. As such, military operations targeting private data victimize civilians during hostilities and thus infringe one of the core objectives of IHL.

Nonetheless, operations to access, copy, or even publish private data will always remain significantly below the threshold of "acts or threats of violence the primary purpose of which is to spread terror among the civilian population" (Art. 51(2) API). Whether the non-violent and non-destructive access to data would really be a "most serious crime of international concern" (Art. 1 RS) is highly questionable. So far, espionage of civilians, as infringing on human rights as it may be, has equally not been considered a war crime (Rule 107). Such acts seem generally better governed by transnational criminal law (e.g., Art. 2 Budapest Convention). A combination of extensive interpretation of the RS, general international law, and national criminal law seems sufficient to govern currently conceivable cyber operations. Ultimately, hackers are not free from international criminal accountability and death and destruction of civilian objectives brought about by pressing 'ENTER' can certainly get you to The Hague.

The "Bofaxe" series appears as part of a collaboration between the IFHV and Völkerrechtsblog.
